

ORGANIZAÇÃO DA INFORMAÇÃO E BIBLIOMETRIA

INFORMATION ORGANIZATION AND BIBLIOMETRY

Lígia Café - ligia@cin.ufsc.br
Universidade Federal de Santa Catarina
Marisa Bräscher - marisab@unb.br
Universidade de Brasília

Resumo

Padrões de organização da informação são necessários em pesquisas bibliométricas para que documentos que possuam características similares possam ser recolhidos em fontes primárias e secundárias e para que se possa atingir confiabilidade na análise dos dados. Com o objetivo de identificar as dificuldades encontradas por pesquisadores quanto aos aspectos de organização da informação, efetuou-se um levantamento de 59 relatos de teses e dissertações e realizaram-se entrevistas com especialistas da área. Conclui-se que, de maneira geral, as dificuldades se concentram principalmente na padronização de títulos de periódicos e de nomes de autor (descrição física), assim como na diversidade de representações de assunto (descrição de conteúdo). A necessidade de controle da literatura por meio da manutenção de bases de dados atualizadas é apontada pelos especialistas entrevistados como uma das principais dificuldades encontradas na realização de pesquisas na área de bibliometria no Brasil.

Palavras-chave: Bibliometria. Organização da Informação

1 INTRODUÇÃO

De maneira geral, a Bibliometria pode ser definida como um conjunto de leis e princípios aplicados a métodos estatísticos e matemáticos que visam o mapeamento da produtividade científica de periódicos, autores e representação da informação. Para cada um destes campos de estudos são adotadas respectivamente as leis de Bradford, de Lotka e de Zipf. Outros estudos que examinam, por exemplo, a análise de citações (identificação de frentes de pesquisa e colégios invisíveis, fator de imediatismo ou de impacto, acoplamento bibliográfico e co-citação, obsolescência da literatura e vida-média, lei do elitismo e teoria epidêmica de Goffman) também se inserem no contexto das pesquisas bibliométricas. A grosso modo, estas leis sintetizam-se da seguinte forma: Bradford objetiva conhecer o núcleo de periódicos produzido em determinado tema, Lotka visa definir as maiores contribuições de pesquisadores em determinadas áreas do conhecimento e Zipf pontua a frequência com que certas palavras aparecem nos textos científicos de maneira a definir sua representatividade neste contexto.

As três leis citadas lidam com distribuições e recenseamentos de documentos científicos que possuem propriedades similares. Estas distribuições são interpretadas sob a luz de dois conceitos principais: o núcleo e a dispersão. Segundo Rostaing (1996, 26),

***Núcleo:** representa o grupo de elementos que aparecem mais freqüentemente em um conjunto de referências bibliográficas estudadas. Por exemplo, no caso da lei de Lotka, o núcleo simboliza os autores mais produtivos em determinada área do conhecimento.*

***Dispersão:** representa o número de elementos de baixa freqüência no conjunto de referências bibliográficas estudadas. No caso da lei de Lotka, a dispersão corresponde a uma grande diversidade de autores que publicaram pouco nessa mesma área do conhecimento.*

Para que se possam recolher documentos científicos que contenham propriedades similares, é essencial a padronização da descrição física e de conteúdo desses documentos. Uma organização padronizada da informação em bases de dados proporciona a recuperação de itens relevantes que revelarão uma distribuição mais próxima da realidade e, conseqüentemente, a verificação adequada dos conceitos de núcleo e dispersão.

2 ORGANIZAÇÃO DA INFORMAÇÃO

Ao associar organização e recuperação da informação, Svenonius (2000) enfatiza que a eficácia de um sistema de recuperação da informação está diretamente relacionada à inteligência adotada para sua organização. Nesta direção, a autora afirma que “o objetivo essencial e que define um sistema de organização da informação é o de agrupar informação semelhante e diferenciar aquela que não é exatamente semelhante.” (SVENONIUS, 2000, p.11.)

Shera e Egan (1961) são ainda mais precisos ao definir que o objetivo dos processos de documentação – bibliografias, serviços de resumos, índices e classificação – é a individualização de um determinado item entre o vasto número dos que formam o conjunto de literatura.

A individualização de um item de informação é obtida por meio da descrição física e de conteúdo. Svenonius (2000) lembra que, para ser organizada, a informação precisa ser descrita. Uma descrição, por sua vez, necessita de uma linguagem. Segundo a mesma autora, a descrição bibliográfica utiliza a linguagem bibliográfica, que se desdobra em linguagens que descrevem a informação (conteúdo) e linguagens que descrevem o documento (suporte físico).

Le Coadic (2004) também faz a distinção entre a descrição física e de conteúdo quando se refere aos métodos de análise dos documentos e da informação, ao afirmar que esses métodos “visam a extrair do documento um conjunto de palavras que sirvam para representá-lo de forma condensada. Dois tipos de conjuntos podem ser obtidos: o primeiro remete à origem do documento e o segundo, ao seu conteúdo.” (LE COADIC, 2004, p. 64)

Nas definições citadas, observa-se que a descrição é o elemento chave da organização da informação. Na recuperação da informação, a descrição é também fundamental,

pois cumpre a função de possibilitar o acesso a itens específicos que se encontram armazenados.

Da mesma forma que para a recuperação da informação, a descrição física e de conteúdo é essencial para os estudos bibliométricos. O produto dessa descrição é a individualização de um item de informação e a identificação das suas características e propriedades. Sem a descrição e, conseqüentemente, sem a identificação única de um item de informação, não é possível realizar análises bibliométricas, pois, como ressalta Alvarenga (1998), os “elementos textuais, paratextuais e contextuais referentes a monografias e artigos de periódicos se constituem em variáveis comumente abordadas nos estudos bibliométricos.”

Na literatura da área de Ciência da Informação é consagrado que os sistemas de recuperação da informação devem ser desenhados em função de seus usuários. Geralmente, quando se organiza informação, se tem em vista o usuário final de um sistema de informação, que recorre ao sistema para satisfazer uma necessidade específica de informação. No entanto, os pesquisadores que desenvolvem estudos bibliométricos também necessitam de um conjunto de descrições padronizadas para constituir seu *corpus* de pesquisa e realizar suas análises.

Neste artigo, procura-se mostrar a importância da organização de informação para os usuários que necessitam recuperar e individualizar itens de informação para a realização de estudos bibliométricos. O objetivo do artigo é identificar dificuldades no que se refere à descrição física e de conteúdo das fontes utilizadas nesses estudos.

3 METODOLOGIA

Para colher evidências quanto à existência de dificuldades relativas à ausência de padronização da descrição bibliográfica na realização de estudos bibliométricos, e, ainda, com o objetivo de identificar a natureza dessas dificuldades, realizou-se um levantamento de relatos de pesquisas de mestrado e doutorado na área de Bibliometria.

Analisaram-se 49 teses e dissertações obtidas como resultado de busca realizada na base de dados de literatura em Ciência da Informação - LICI, do IBICT. A estratégia de busca utilizada foi o descritor *bibliometria* no campo assunto e *dissertação e tese* no campo tipo de documento da base de dados.

Visando a análise de pesquisas mais recentes do que aquelas recuperadas na LICI, consultou-se a Biblioteca Digital de Teses e Dissertações (BDTD)¹ do IBICT utilizando-se a palavra-chave *bibliometria*. Dos 17 registros recuperados, quatro já haviam sido recuperados na LICI, dois apresentaram erro de *link* que impediu o acesso ao documento integral e um registro estava duplicado. Da BDTD restaram, portanto, 10 documentos a serem analisados. O conjunto de documentos recuperados nas duas fontes citadas encontra-se referenciado ao final deste artigo.

¹ <http://bdttd.ibict.br> (acesso em 05/09/2007)

Procedeu-se à leitura dos itens metodologia, universo de coleta de dados, análise dos dados e conclusões de cada tese e dissertação recuperada na LICI e na BDTD. Realizou-se o registro e agrupamento das dificuldades segundo sua natureza, isto é, se eram relativos a problemas de descrição física ou de descrição de conteúdo.

Este levantamento não teve a intenção de ser exaustivo e tampouco de apresentar dados quantitativos relacionados às dificuldades levantadas pelos pesquisadores. As informações obtidas possibilitaram uma análise qualitativa das dificuldades relatadas e sua caracterização quanto aos aspectos de organização da informação.

Ainda como parte da metodologia, realizou-se entrevista com 4 pesquisadores brasileiros² da área de Bibliometria, com o objetivo de identificar se encontravam ou não dificuldades relativas à descrição física e de conteúdo nas fontes utilizadas nos estudos bibliométricos e, em caso afirmativo, de obter o relato das dificuldades encontradas.

4 A DESCRIÇÃO BIBLIOGRÁFICA E OS ESTUDOS BIBLIOMÉTRICOS

Das 59 teses e dissertações analisadas, 30 apontaram algum tipo de problema quanto à descrição física ou de conteúdo das fontes utilizadas para coleta de dados. Não se pode afirmar que a ausência de relatos nos demais trabalhos significa que os autores dessas pesquisas não encontraram dificuldades, talvez apenas não as tenham mencionado.

Apesar de extraído de um universo reduzido de trabalhos de pesquisa, este dado fornece um indício de que, em pesquisas da área de Bibliometria, ocorrem efetivamente problemas oriundos da ausência de padronização de descrições bibliográficas.

a. Problemas de descrição física

A qualidade da descrição física nas bases de dados afeta diretamente a recuperação para efeito de pesquisa bibliométrica, principalmente quando se lida com as leis de Bradford e de Lotka. Isto porque a primeira, ao procurar verificar a capacidade de um periódico em representar uma determinada área do conhecimento, utiliza a recuperação por título dos artigos e periódicos. Quanto à lei de Lotka, a falta de padronização da descrição física afeta principalmente a recuperação dos nomes de autores, uma vez que esta lei procura medir a produtividade científica dos pesquisadores como elemento de representatividade nas diversas áreas científicas.

Rostaing (1996, p.28) aponta as seguintes questões relacionadas ao tema.

- a) *Quanto à avaliação de revistas, o maior problema relaciona-se ao uso de uma diversidade de abreviações no título;*
- b) *Quanto à avaliação de pesquisadores (apud MCROBERTS, 1989):
. somente o primeiro autor do artigo citado é considerado*

² Agradecemos a valiosa colaboração dos professores pesquisadores Gilda Braga, Jaime Robredo, Lena Vânia Ribeiro Pinheiro e Raimundo Nonato.

- . *risco de homonímia*
- . *risco de auto-citação*
- . *erros ortográficos.*

Para a análise de citações é necessário, portanto, identificar de forma única um título e um autor quando este aparece nas diversas citações analisadas. Caso não esteja padronizado, um mesmo título ou um mesmo autor poderá ser contabilizado como diferente.

Quanto a este aspecto, foi possível registrar que a ausência de padronização dos títulos de periódicos e nomes de autores apresenta-se como uma dificuldade para os estudos que trabalham com análise de citações, como exemplificado em alguns relatos descritos a seguir.

Cavalcante (1989) aponta a necessidade de uniformizar e completar os dados das referências bibliográficas, pois algumas bibliografias não seguiam os padrões estabelecidos.

Lopes (1992) indica que grande parte das citações não trazia local de edição, havendo necessidade de consulta a fontes de referência. Quanto aos títulos de periódicos, observou que “Ora grande parte dos títulos vinham abreviados de forma incorreta, ora traziam as nomes grafados de forma errônea, com as palavras trocadas de lugar.” (LOPES, 1992, p.60)

Decourt (1993) relata que uma etapa importante da pesquisa, após a seleção de artigos a serem incluídos, foi a uniformização dos nomes dos autores e a codificação dos títulos de periódicos. As regras do AACR2 foram aplicadas a fim de uniformizar as entradas e, assim, evitar a ocorrência de homônimos.

Para atingir o objetivo de recuperar toda a produção científica de um mesmo autor, Fernandes (1973) ressalta que, em seu estudo, houve necessidade de investigar o nome completo dos autores e supor todas as possíveis entradas para um mesmo autor nas fontes primária e secundária.

Costa (2007) observa a falta de padronização das referências bibliográficas, “títulos com abreviações diferentes, referências colocadas ora no corpo do documento, ora no final”. Magalhães (2006) menciona a omissão e troca de dados em referências bibliográficas e Leal (2005) refere-se a autores que apareciam mais de uma vez, nomes escritos de forma diferente e assuntos com grafia diferente que remetem ao mesmo conceito.

Acerca da padronização do nome do autor, além dos exemplos mencionados, a ausência de dados completos relativos a outras características da autoria como a titulação e a afiliação, por exemplo, também se encontra relatada nos estudos.

Souza (1982) aponta que a insuficiência de dados sobre os professores/pesquisadores da instituição objeto de sua pesquisa dificultou o estudo e afirma que as informações referentes à titulação, por exemplo, enriqueceriam as análises.

Segundo Lopes (1974), os autores do periódico analisado em sua pesquisa não traziam dados completos sobre titulação e local de graduação e, portanto, houve necessidade de completar esses dados em consulta à área de recursos humanos das respectivas instituições de origem.

Carvalho (2006) chama a atenção para a inclusão, a partir de 2002, de um campo específico na base de dados Bibliografia Brasileira de Odontologia (BBO) para cadastramento da afiliação do autor. Por outro lado, a ausência de dados da afiliação de autores secundários no Medline é apontada como uma dificuldade pela autora. Dos 90 autores analisados em sua pesquisa, 21 não tiveram esse dado preenchido.

Diante dos problemas encontrados, alguns relatos enfatizam a importância da padronização da descrição física dos documentos para as análises de citação. Nesse aspecto, Souza (1992) destaca a necessidade de aplicação das normas de descrição:

Merecem atenção especial os problemas que prejudicam a coleta de dados da literatura citada e que são resultantes da falta de normalização das citações dentro dos padrões estabelecidos pelos órgãos normativos, nesse caso a NBR 6023 da ABNT que, ao que parece, não é exigida, o que impossibilita e dificulta o cotejamento entre as informações citantes e as informações citadas. (SOUZA, 1992, p.78)

Em sua pesquisa de mestrado, a visão de Braga (1977, p.16), como uma das grandes especialistas na área de Bibliometria no Brasil, resume a problemática:

As formas de citar, embora indicadas na maioria dos periódicos, não são, freqüentemente, normalizadas, completas e precisas: citações com mais de um autor têm o segundo autor e demais autores substituídos por et alii, et al. etc. ou omitidos; as indicações de volume, fascículo e data dos periódicos variam enormemente, sendo muitas vezes errôneas.

Outro aspecto levantado nos estudos analisados refere-se à ausência de padronização na estrutura dos periódicos, como indicado por Pinheiro (1982), que identifica o problema em diferentes fascículos de uma mesma revista. Este fato constitui-se em obstáculo para identificar, nos periódicos, quais seções devem ser consideradas para levantamento dos artigos científicos e de suas citações.

Essa mesma dificuldade encontra-se relatada em estudo recente sobre o periódico Ciência da Informação, realizado por Pinheiro; Bräscher e Burnier (2005, p.33), que identifica 22 tipos de seções na revista. Ressaltam as autoras que

...não significa que o periódico não passe por nenhuma modificação, ao longo dos anos, nem que possa ser criada uma nova seção, mas o desejável é que mantenha um padrão estrutural e terminológico que, inclusive, identifique-o.

Nesse estudo de Pinheiro; Bräscher e Burnier (2005), os dados relativos à autoria, revelam dificuldades semelhantes às relatadas nas teses e dissertações levantadas neste artigo. Das 856 ocorrências de autores da revista, 42,52% não indicava o grau de formação dos autores. Além disso, os nomes de autores e de suas instituições de origem também tiveram que ser padronizados para possibilitar a contagem de ocorrência de autoria e para identificar as instituições de origem que mais contribuíram com a revista.

Nos estudos analisados, as deficiências de padronização da descrição física são apontadas como uma dificuldade pelos pesquisadores. A ausência dessa padronização exige um esforço adicional dos autores das pesquisas, que necessitam recorrer a várias fontes de consulta para completar e confirmar dados necessários às análises bibliométricas.

Como ressalta Mugnaini (2006, p.), “o uso de referências bibliográficas para análises bibliométricas impõe a transformação de dados bibliográficos em dados bibliométricos. Desta forma, a padronização da informação é uma exigência, já que se deve garantir a homogeneidade das informações a serem compatibilizadas”. Um dado interessante relatado em seu estudo mostra que, após a padronização do título, determinada revista que ocupava a primeira posição no ranking passou a ocupar o sexto lugar.

Mugnaini (2006, p.), apoiado na opinião de diversos autores, chama a atenção para o fato de, no Brasil, muitos estudos se limitarem aos dados fornecidos no ISI (Information Sciences Institute) em função de “grande parte do esforço de realização das análises bibliométricas centrar-se na padronização dos dados, quase sempre obtidos de bases de dados bibliográficas onde a falta de uniformidade é extrema”.

b. Problemas de descrição de conteúdo

Um dos aspectos relativos à descrição de conteúdo diz respeito à coleta, em bases de dados, bibliografias ou catálogos, das publicações dedicadas à área de conhecimento objeto do estudo bibliométrico e que compõem o corpus de pesquisa.

Nesse aspecto, as origens das dificuldades para os estudos bibliométricos são semelhantes às aquelas encontradas em sistemas de recuperação da informação. Isto se deve ao fato de que a descrição e a recuperação de conteúdo supõem o cruzamento das noções de especificidade e exaustividade, assim como de relevância, revocação e precisão.

Segundo o princípio da especificidade, os assuntos devem ser representados da forma mais específica possível. Assim, como exemplifica Lancaster (2004), um documento

sobre "laranja" será descrito pela etiqueta "laranja" e não por "fruta cítrica" ou "fruta". A exaustividade refere-se ao nível de profundidade dado à descrição de conteúdo de um documento. De acordo com Lancaster (2004), a indexação exaustiva abrange o conteúdo temático do documento de modo bastante completo e a indexação seletiva abrange somente o conteúdo temático principal do documento. Estes princípios visam melhorar a relevância, segundo as necessidades de informação dos usuários de um sistema de informação.

Como afirma Le Coadic (2004), a relevância é uma relação entre um documento e uma questão do usuário. Dessa forma, Lancaster (2004) define que um documento pertinente (útil) é aquele que contribui para satisfazer alguma necessidade de informação. O conceito de relevância se relaciona, portanto, com os significados de revocação e precisão, medidas de avaliação de sistemas de recuperação da informação.

A revocação é a relação entre o número total de documentos relevantes contidos no sistema de informação e o número de documentos relevantes recuperados pelo sistema. A precisão, por sua vez, é a relação entre o número de documentos relevantes recuperados pelo sistema e o número total de documentos recuperados. A precisão, portanto, diz respeito à capacidade do sistema de recuperar apenas os documentos relevantes, ou seja, sua capacidade de descartar documentos inúteis e assim diminuir o ruído na recuperação da informação. (ROBREDO, 2005).

Attingir índices adequados de revocação e de precisão é uma condição importante para a recuperação de informação em pesquisas bibliométricas, uma vez que o conjunto de documentos recuperados que constitui o *corpus* de pesquisa afeta diretamente o resultado das análises de distribuição realizadas.

A articulação dos conceitos de especificidade, exaustividade, revocação e precisão deve ter como orientação maior a relevância do assunto para aquele que está buscando a informação, o que nem sempre é evidente. Nesse sentido, a adoção de linguagens documentárias pode auxiliar os processos de representação e recuperação da informação. É muito importante salientar que as linguagens documentárias, ao visarem à padronização, não devem representar a terminologia de uma área do conhecimento de forma prescritiva como se as línguas de especialidades pudessem ser controladas. É necessário padronizar respeitando a diversidade denominativa e nocional da área descrita, levando em conta o uso social da língua para a comunidade à qual se destina. Esse aspecto, no entanto, pode ocasionar alterações na estrutura da linguagem, o que vem muitas vezes dispersar a informação representada, dificultando o trabalho bibliométrico.

Com relação ao uso de linguagens documentárias, os relatos encontrados nas teses e dissertações referem-se à ausência de padronização na representação dos assuntos; às alterações na estrutura da linguagem documentária utilizada ou às deficiências da linguagem. Nesse sentido, o retrabalho ou a adoção de remissivas nas bases de dados é muito importante para que a determinação dos recenseamentos seja a mais próxima da realidade.

Dentre os estudos analisados, Rolim (1976, p.63) ressalta a importância da padronização de assunto ao relatar que “Com os dados obtidos e organizados, a primeira preocupação foi a normalização de cabeçalhos de assunto sob números idênticos de classificação, a fim de se efetuar o tratamento estatístico”. A pesquisadora aponta, ainda, dificuldades com a alteração na estrutura da linguagem documentária utilizada pela bibliografia analisada em seu estudo.

Obviamente, não seria possível distinguir, sem acesso às publicações, qual a razão de se encontrarem reunidos assuntos como Arquitetura e Escultura, que aparecem juntos em 1952, embora depois passem a ser indicados isoladamente.

Constata-se ainda que o assunto poesia e poemas musicais a partir do ano XXI, 1960, passou a figurar sob o título POEMAS e que o assunto CINEMA foi omitido, a partir de 1959, ano XX, da rubrica PEÇAS DE TEATRO, RADIOTEATRO, CINEMA E TELEVISÃO.” (ROLIM, 1976, p.64)

A necessidade de padronização da descrição é reconhecida também por Fernandes (1973), ao analisar a literatura de Física, quando afirma ter sido necessário, a fim de obter todos os dados classificados de uma mesma maneira, escolher uma lista classificada de cabeçalhos de assunto.

Do relato de Coutinho (1990), é possível identificar como os problemas de indexação afetam os estudos. A autora observa que artigos sobre ferrovia – tema de sua pesquisa - deixaram de ser indexados na base de dados estudada, por se referirem a assuntos mais complexos e, portanto, com maior grau de dificuldade para identificação do assunto tratado e, conseqüentemente, mais difíceis de atribuir cabeçalhos de assunto satisfatórios. A autora atribui esses problemas de indexação à excessiva especialização do assunto e à existência de um tesouro especializado em ferrovia que, possivelmente, não estava sendo usado adequadamente ou não atendia às necessidades de indexação.

Por outro lado, alguns relatos apontam que descrições padronizadas de assunto contribuem para a coleta de dados em estudos bibliométricos. Pinheiro (1982, p.43) utiliza a base LISA (Library Information Science Abstract) em seu estudo bibliométrico e não encontra dificuldades ao selecionar os registros relativos ao tema de sua pesquisa – classificação, pois, como afirma, “essa área mostrou ser de fácil recuperação porque os descritores têm seus limites fixados sem o uso excessivo de remissivas e referências.”

A base de dados LISA utiliza linguagem documentária para a descrição dos assuntos. Dessa maneira, o pesquisador da área de Bibliometria, assim como um usuário que deseja recuperar informação nessa base de dados, é orientado a selecionar os descritores que melhor representam os assuntos de seu interesse.

Faria (2001) menciona as facilidades com o uso de descritores e modificadores do tesauro para obter maior precisão na busca no Web of Science e, por outro lado, aponta como dificuldade a ausência de um campo de classificação por áreas do conhecimento para permitir a recuperação de todos os artigos de uma área, de forma mais abrangente do que a possibilitada pelos descritores específicos.

Os relatos de pesquisa revelam que a qualidade da descrição de conteúdo afeta a qualidade da recuperação da informação e também os estudos bibliométricos que dependem dessa descrição para delimitar seu *corpus* de pesquisa.

Além dos fatores ligados à recuperação de documentos por assunto em bases de dados, outro tipo de dificuldade apontada nos estudos analisados diz respeito à aplicação da lei de Zipf, conforme ressaltam os relatos citados a seguir.

Como aponta Silva (1995), a lei de Zipf trabalha com palavras tratadas isoladamente e os resultados obtidos poderiam ser significativamente alterados utilizando-se outros critérios de contagem de palavras ou adotando-se conceitos. O autor ressalta ainda a necessidade de considerar o problema dos quase-sinônimos, apontando como exemplo os termos soropositivo, AIDS, HIV.

No entanto, ao considerar os quase-sinônimos em um mesmo conjunto de dados recuperados sem levar em consideração suas proximidades e naturezas semânticas e pragmáticas, pode-se chegar a resultados errôneos. Os conceitos citados acima, mesmo que relacionados, têm significados diferentes dependendo do contexto no qual estão inseridos. Isso deve ser levado em consideração no momento da análise dos dados recuperados.

Maranhão (1996), ao aplicar a lei de Zipf em dois textos, um em português, e sua tradução para o inglês, e vice-versa, confirma o que Silva infere, pois ao aplicar a fórmula de transição de Goffman ao conjunto de palavras extraído do texto em inglês e, posteriormente, ao conjunto extraído da sua tradução para o português, observa que o termo *information* encontra-se na faixa de palavras representativas do conteúdo, ou seja, aquelas que se encontram no ponto T da fórmula de transição de Goffman (usa a variação da fórmula conforme proposto por Mamfrim). Ao aplicar a lei no texto em português, o termo correspondente em português 'informação' não é incluído entre as palavras do ponto T.

A autora observa que isso se deve ao fato de a palavra inglesa não possuir flexão de número enquanto que na língua portuguesa ela sofre flexão e, caso não ocorra redução ao radical, as frequências das duas formas são contadas distintamente, levando a diferenças significativas de resultados.

Charles Muller *apud* Quemel (1983) chama a atenção para a necessidade de regras precisas para a delimitação da palavra, unidade de texto e de vocábulo, unidade do léxico:

quantificar o vocabulário de um texto é proceder a duas operações distintas, sejam elas sucessivas ou simultâneas:

1) a enumeração das palavras que compõem o texto, e cujo número representado por N, dá uma medida de extensão do texto;

2) a enumeração dos vocábulos empregados no texto, e cujo número, representado por V, mede a extensão do vocabulário.

É necessário, então, regras precisas para a delimitação da palavra, unidade do texto, e de vocábulo, unidade de léxico.” (MULLER, apud QUEMEL, 1983, p.31)

Como conclui Quemel (1983), “as leis de Zipf são referentes à extensão do texto, ou seja, a verificação é feita pela enumeração das palavras e não de vocábulos.” Dessa forma, “são computados separadamente os sinônimos, palavras no singular e plural, diversos tempos do mesmo verbo, palavras de gênero diferentes, enfim, todas as flexões de um vocábulo.”

A análise da autora se refere à diferença entre duas abordagens de tratamento textual presentes na literatura: a abordagem estatística e a lingüística. A primeira faz uso de recursos estatísticos e se baseia na freqüência de ocorrência das palavras no texto e na base de dados analisada. Apesar das vantagens de serem mais rápidas e necessitarem de menores recursos de informática e de conhecimento, estudos em Lingüística têm apontado as seguintes desvantagens em relação a esta abordagem:

- a) dificuldade em estabelecer uma medida ideal de freqüência para medir a representatividade da palavra no contexto estudado;
- b) falta de controle de ambigüidades semânticas;
- c) a recuperação é afetada negativamente no que diz respeito aos níveis de precisão e revocação.

Por outro lado, a abordagem lingüística se fundamenta em técnicas mais sofisticadas, chegando por vezes a adotar recursos de manipulação da linguagem natural chamados de processamento automático da linguagem natural. A descrição de conteúdo de abordagem conceitual pode resolver ambigüidades morfológicas, lexicais, sintáticas e semânticas. Em contrapartida, obtém-se um conjunto mais significativo de itens de dados para o trabalho bibliométrico, aproximando-o da realidade.

Basílio, *apud* Ramos (1992), defende a abordagem lingüística ao ressaltar a necessidade de considerar que

a contagem das palavras de um texto deve ser feita não em termos das palavras propriamente ditas, e sim em termos do que ela chama de ‘contemas’, ou seja, de unidades de conteúdo informativo, pois como a língua portuguesa é rica em palavras de forma gráfica e função sintática diferentes, mas equivalentes quanto ao conteúdo informativo, uma contagem que separasse estas diversas formas de um mesmo conteúdo produziria resultados distorcidos. (RAMOS, 1992, p.320)

Em estudos que utilizam métodos estatísticos de tratamento as palavras do texto, como no caso dos que aplicam a Lei de Zipf, os resultados podem ser significativamente alterados dependendo do tipo de tratamento lingüístico aplicado ao texto. É necessário, portanto, que os procedimentos adotados na pesquisa estejam bem especificados na metodologia, como foi o caso de várias teses e dissertações analisadas, que apresentam as regras adotadas para o tratamento textual, tais como a redução das palavras ao radical e o tratamento de termos compostos e de nomes próprios.

Quanto à descrição de conteúdo, a análise das dissertações e teses revela que os estudos bibliométricos são afetados por fatores lingüísticos da linguagem natural semelhantes aos que afetam o processo de indexação e de elaboração de linguagens documentárias, tais como a sinonímia, a homografia e o tratamento de termos compostos.

5 A OPINIÃO DOS ESPECIALISTAS DA ÁREA DE BIBLIOMETRIA NO BRASIL

Nas entrevistas com especialistas brasileiros da área de Bibliometria, algumas opiniões corroboram os relatos encontrados nas teses e dissertações analisadas quanto à necessidade de padronização dos títulos de periódicos e de nomes de autores. Ressalta-se o risco de se contar o mesmo periódico como periódicos distintos, quando a fonte adotada para coleta dos dados não recupera os diferentes títulos, fusões etc. Em relação à autoria, além do aspecto da padronização dos nomes, um dos especialistas lembra ainda que “as abreviaturas de nomes podem também induzir a erros quando, coincidentemente, o mesmo sobrenome também apresenta as mesmas letras no pré-nome”. Nesse caso, como alerta o pesquisador, somente especialistas da área estudada são capazes de identificar essas idiosincrasias.

A visão de um segundo especialista entrevistado confirma que os problemas se agravam quando os estudos são realizados em áreas que o pesquisador desconhece, pela dificuldade de reconhecer títulos e autores. Esse especialista afirma que “o trabalho bibliométrico elaborado na própria área é menos árduo, pois conhecemos o que estamos analisando. O problema aparece quando vamos para outras áreas do conhecimento.”

Outro aspecto relativo a autoria diz respeito à orientação de uso do *et al* dada pela NBR 6023 da ABNT (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2002), identificada como uma das dificuldades encontrada pelos especialistas, pois, a ausência do nome dos co-autores inviabiliza estudos de produtividade ou de citação.

Além dos problemas de títulos e nomes e autores, um dos especialistas aponta ainda a falta de padronização do nome de instituições. Uma breve consulta ao Web of Science³ para levantar a produção oriunda da Fundação Oswaldo Cruz (Fiocruz), por exemplo, revela a falta de padronização no nome de instituições brasileiras. Com o nome completo da instituição recuperaram-se 3564 registros e com a sigla 3.329.

³ <http://portal.isiknowledge.com/portal.cgi?DestApp=WOS&Func=Frame> (acesso em 10 de setembro de 2007)

Pôde-se observar nessa rápida consulta, que o mesmo ocorre com outras instituições pesquisadas, como a Universidade de Brasília, que ora aparece abreviada - Univ Brasília, e ora é representada pela sigla UNB. O levantamento de dados de instituições requer, portanto, que o pesquisador imagine todas as possibilidades de registro do nome da instituição que deseja estudar.

Segundo um dos entrevistados, com a aplicação de ferramentas de processamento automático de texto é possível formar agrupamentos de nomes semelhantes, o que facilita a identificação de autores, títulos e instituições e sua contagem única nas análises bibliométricas. Esse mesmo especialista ressalta que os agrupamentos se comportam de maneira diferente em áreas do conhecimento distintas.

Uma solução descrita na tese de Mugnaini (2206) também aponta na direção do uso de ferramentas para correção automática de diversos campos, como o módulo de bibliometria do Scielo, utilizado em sua pesquisa, que trabalha com associações por trigramas e buscas booleanas. No entanto, o próprio autor reconhece que este processo de correção pode falhar em função da insuficiência de dados nas fontes utilizadas.

Sobre a descrição de conteúdo, apenas um especialista se manifestou e identificou o seguinte problema em relação à Lei de Zipf:

O maior problema atual, em aplicações da lei de Zipf, é encontrar software apropriado para operar como os conteúdos semânticos (algumas vezes palavras compostas), enfim, um trabalho intelectual realizado bem por um ser humano. O trabalho, quando manual, exigia um grande esforço para este tipo de detalhamento.

No contexto deste artigo, as características de softwares aplicados aos estudos bibliométricos não foram analisadas. Uma sugestão para estudo futuro é verificar o quanto esses softwares são capazes de resolver os aspectos mencionados pelo especialista.

Sobre as questões relativas à padronização dos dados, um dos especialistas opina que estes não são o maior problema das pesquisas bibliométricas, mas sim a carência de fontes secundárias. Segundo ele,

as principais dificuldades não se referem à aplicação de determinadas leis – como Bradford, Zipf ou Lotka, mas sim à existência de fontes secundárias adequadas – e quando estas existem, acesso às fontes primárias. Padronização é luxo – e não incomoda muito. Um certo ruído nos dados é inevitável, como em qualquer levantamento ou pesquisa sobre esses dados – um pesquisador deve saber lidar com isso.

Três dos especialistas entrevistados concordam que a maior dificuldade para os estudos refere-se à necessidade de controle da literatura, principalmente por meio da manutenção de bases de dados atualizadas da literatura brasileira.

Até a década de 60, o IBICT compilava bibliografias manualmente – depois automatizadas a duras penas, ainda via cartões perfurados. Hoje em dia, com todo o poder de computação disponível, a literatura científica e tecnológica brasileira é uma grande incógnita.

Sem bases de dados com padrões, completas e atualizadas, não há possibilidade de obter indicadores de C&T confiáveis e, conseqüentemente, de planejamento de Ciência e Tecnologia, com base nesses indicadores.

Muitas das dificuldades de trabalhar com bibliometria no Brasil devem-se a deficiências das bases de dados nacionais.

Sobre o uso de bases de dados como fontes de coleta de dados em pesquisas da área de Bibliometria, um especialista ressalta que um dos problemas encontrados é que essas geralmente não são construídas para essa finalidade e, portanto, há dificuldade na coleta e padronização de dados.

Um apontamento interessante feito por um dos entrevistados refere-se ao endereçamento das sugestões aos órgãos do governo responsáveis pelas políticas públicas em informação científica e tecnológica – ICT, “sobretudo o IBICT, que é a instituição brasileira com esta competência, não somente de orientação para a adoção de padrões nacionais, compatíveis com os internacionais, mas de apoio a essas iniciativas, evitando a situação de descontinuidade e desatualização, que é muito comum no Brasil.”

Quanto ao uso de *et al* na normalização da referência bibliográfica pela ABNT, um dos entrevistados propõe que esta instituição seja orientada sobre estas questões, que atingem diretamente estudos quantitativos como os bibliométricos.

Finalmente, outra questão citada nas entrevistas diz respeito à importância do levantamento de teses e dissertações com abordagem bibliométrica, uma vez que os procedimentos metodológicos, de um modo geral, são muito detalhados, exatamente para superar esses problemas apontados chegando, às vezes, a constituir capítulos. Sobre este assunto, um entrevistado cita uma pesquisa que embora antiga é ainda hoje atual quanto ao registro da problemática. Diz o entrevistado

É oportuno lembrar a dissertação de mestrado de Helena Dodd Ferrez, na área de História que, ao enfrentar tantos problemas de citação, inseriu na dissertação um capítulo intitulado “patologia de citações”. Mesmo que esta dissertação seja do início da década de 80, passados mais de 20 anos ainda há muitos problemas que persistem e permanecem.

A consulta à dissertação de Ferrez (1981) leva à compreensão da ênfase dada às “patologias de citações”. Pois, em seu estudo de citações na área de História, a autora encontrou 75% de referências incompletas, o que a motivou a escrever um capítulo específico sobre esse problema (FERREZ, 1981). Dentre as possíveis causas dessa ausência de padronização e completeza das citações, a pesquisadora ressalta o grau de maturidade em que se encontrava a área de História à época; a inexistência de um corpo editorial que efetivamente opere a partir de normas pré-determinadas, que assegurem um padrão qualitativo homogêneo; a complexidade das referências da área e o desconhecimento da importância da referência no sistema de comunicação científica.

Com base na análise de teses e dissertações realizada neste artigo, pode-se afirmar que problemas de descrição bibliográfica continuam a ser apontados em estudos bibliométricos, mesmo nos mais recentes, apesar de nenhum relato se comparar à situação enfrentada por Ferrez (1981).

6 CONCLUSÕES

Ainda que preliminarmente, o estudo realizado revela que questões relativas à padronização da descrição bibliográfica causam dificuldades na coleta e análise de dados em pesquisas bibliométricas.

No levantamento e análise realizados constataram-se dificuldades dos estudos de análise de citação no que se refere à descrição bibliográfica, mais especificamente, aquelas relativas à padronização de títulos de periódicos e nomes de autores. Os problemas de descrição de conteúdo dizem respeito à representação de assunto nas bases de dados ou outras fontes utilizadas para a coleta dos documentos que compõem o corpus de pesquisa. Na aplicação da Lei de Zipf, há ainda questões relacionadas às características da linguagem natural que necessitam de tratamento lingüístico, tais como variação de número, sinonímia, homografia e nomes compostos.

Tomando por base algumas sugestões de autores de teses e dissertações e de especialistas entrevistados, enumeramos abaixo algumas iniciativas que podem contribuir com a coleta e análise de dados em pesquisas bibliométricas.

- Autores: observar a legitimidade e fazer corretamente a transcrição dos dados nas citações, segundo as normas estabelecidas.
- Editores de periódicos científicos: definição de padrão para identificação dos nomes de autores e exigência de dados completos de filiação e titulação. Definição de normas de citação e controle de qualidade na sua aplicação em artigos submetidos à revista sob sua responsabilidade.
- Bases de dados: além da padronização relativa aos dados de descrição física, a necessidade de utilização de linguagens padronizadas para descrição de conteúdo, visando auxiliar no levantamento da literatura de uma determinada área e na análise de assunto de artigos de periódicos.

Padrões de organização da informação são essenciais para a identificação correta das variáveis analisadas em pesquisas bibliométricas. No entanto, não basta que esses padrões estejam definidos para atingir os objetivos de recuperação da informação e de geração de indicadores de comunicação científica. Eles necessitam ser efetivamente empregados nas fontes de informação. Para tanto, faz-se necessária atuação firme de editores científicos e administradores de bases de dados e a colaboração e sensibilização dos autores para aplicação correta dos padrões de descrição bibliográfica. Nessa direção, gestões junto a órgãos de fomento e órgãos de normalização visando à organização da informação e ao controle da literatura nacional contribuiriam para a produção de análises de indicadores da produção científica brasileira.

REFERÊNCIAS

ALVARENGA, L. Bibliometria e arqueologia do saber de Michel Foucault: traços de identidade teórico-metodológica. **Ci. Inf.**, Brasília, v.27 n.3, set. 1998. Disponível em <http://www.scielo.br/scielo.php?pid=S0100-19651998000300002&script=sci_arttext&tlng=pt>. Acesso em: 06/09/2007.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR6023**: Informação e documentação: referências: elaboração. Procedimentos. Rio de Janeiro, 2002. 22p.

FERREZ, H.D. **Análise da literatura periódica brasileira na área de História**. 1981. 168 p. Dissertação (Mestrado em Ciência da Informação)- Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1981.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. Tradução Antônio Agenor Briquet de Lemos. 2. ed. rev. e aum. Brasília: Briquet de Lemos/Livros, 2004. 452 p.

LE COADIC, Y-F. **A Ciência da Informação**. 2. ed. Brasília: Briquet de Lemos Livros, 2004. 124 p.

PINHEIRO, L.V.R.; BRÄSCHER, M.; BURNIER,S. **Ciência da Informação: 32 anos (1972-2004) no caminho da história e horizontes de um periódico científico brasileiro**. **Ciência da Informação**, Brasília v.34, n. 3, p.1-52. set./dez. 2005.

ROBREDO, J. **Documentação de hoje e de Amanhã**: uma abordagem revisitada e contemporânea da Ciência da Informação e de suas aplicações biblioteconômicas, documentárias, arquivísticas e museológicas. 4.ed.rev. e ampl. Brasília, Edição de autor, 2005. 410p.

ROSTAING, Hervé. **La bibliométrie et ses techniques**.. Toulouse: Sciences de la Société; Marseille: Centre de Recherche Rétrospective de Marseille, 1996. 131 p (Outils et méthodes, 38).

SHERA, J.H.; EGAN, M.E. Exame do estado atual da Biblioteconomia e da Documentação. In: BRADFORD, S.C. **Documentação**. Rio de Janeiro, Fundo de Cultura, 1961. p. 15-61.

SVENONIUS, E. **The Intellectual Foundation of Information Organization**. 2000. 255p.

REFERÊNCIAS DAS TESES E DISSERTAÇÕES COLETADAS

- AGUIARI, Carmen Sylvia Arantes Leal. **Lista básica de periódicos para o curso de graduação em enfermagem e obstetrícia no Brasil: estudo bibliométrico, 1966-1976.** 1980. 132 p. Dissertação (Mestrado em Ciências da Comunicação)-Universidade de São Paulo, São Paulo, 1980.
- AMATUZZI, Maria Luiza Lotumulo. **Avaliação metodológica dos artigos publicados na área de ortopedia e traumatologia nos anos de 2004 e 2005.** 2007. 93 p. Tese (Doutorado em Ciências)-Universidade de São Paulo, São Paulo, 2007.
- BARRETO, Maria de Fátima. **Análise bibliométrica da estrutura e o padrão da produção bibliográfica dos professores dos Departamentos de Ciências Fisiológicas e Física da UFES, 1984-1987.** 1990. 139. p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1990.
- BIASOTTI, Miriam Mara Dantur de la Rocha. **El uso de la informacion documental: concepto y criterios para su medicion, en el ambito de los sistemas de informacion.** 1994. 310 p. Dissertação (Mestrado em Ciência da Informação)-Univ. Complutense de Madrid, Madrid, 1994.
- BRAGA, Gilda Maria. **Dynamics of scientific communication: an application to science funding policy.** 1977. 76 p. Tese (Doutorado em Ciência da Informação)-Case Western Reserve University, Columbus, 1977.
- BRAGA, Gilda Maria. **Relações bibliométricas entre a frente de pesquisa (research front) e revisões da literatura (reviews): estudo aplicado a ciência da informação.** 1972. 37 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1972.
- BRASIL, Maria Irene. **Estruturas bibliométricas e fontes historiográficas do setor de história da Fundação Casa de Rui Barbosa: um estudo de caso.** 1992. 210 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1992.
- CALDEIRA, Paulo da Terra. **Crescimento da literatura brasileira de doença de chagas: análise bibliométrica.** 1974. 61 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Bibliografia e Documentação, Rio de Janeiro, 1974.
- CARVALHO, Maria Martha de. **Análises bibliométricas da literatura de química no Brasil.** 1975. 71 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Bibliografia e Documentação, Rio de Janeiro, 1975.
- CARVALHO, Telma de. **A produção científica brasileira em odontologia e sua visibilidade nacional e internacional.** 2006. 156 p. Tese (Doutorado em Ciências da Comunicação)- Universidade de São Paulo, São Paulo, 2006.
- CAVALCANTI, Ilce Goncalves Milet. **Padrões de citação em comunicação: análise das dissertações apresentadas a ECO/UFRJ.** 1989. 135 p. Dissertação (Mestrado em

Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1989.

CELIS, Adriana Roa. **A dinâmica da comunidade científica na produção do conhecimento**: um estudo da imunologia no Brasil e na Colômbia. 2002. Tese (Doutorado)- Universidade Estadual de Campinas, Campinas, 2002.

COELHO, Beatriz Amaral de Salles. **Literatura, assunto**: análise de citações em ciência política. 1991. 178. p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1991.

COSTA, Rubenildo Oliveira da. **Análise do uso de periódicos científicos na transição do meio impresso ao eletrônico em dissertações e teses**: o impacto do portal de Periódicos/CAPES na produção do conhecimento. 2007. 143 p. Dissertação (Mestrado em Ciência da Informação)_Universidade Católica de Campinas, Campinas, 2007.

COUTINHO, Eliana. **Aplicação da lei de Bradford a literatura técnica sobre ferrovia**: análise de periódicos e avaliação da base de dados de Rede Ferroviária Federal S.A. 1991. 158 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1991.

DECOURT, Beatriz Aparecida Boselli. **O uso da técnica de citação no estudo dos processos de desenvolvimento e de comunicação de grupos de pesquisa em ciência no Brasil**: um estudo de caso. 1993. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1993.

DI GIORGI, Maria Luiza Andrade. **Análises da comunicação entre autores no campo da literatura brasileira de tecnologia de alimentos**. 1974. 133 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Bibliografia e Documentação, Rio de Janeiro, 1974.

DUSILEK, Darci. **Investigação sobre a influência das variáveis tempo e assunto no vocabulário de um autor**. 1974. 102 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Bibliografia e Documentação, Rio de Janeiro, 1974.

ESPIRITO SANTO, Alexandre do. **A measure of the dimensions of interdisciplinarity of two applied sciences**: a scientometric model. 1978. 136 p. Dissertação (Mestrado em Ciência da Informação)-University of Wisconsin , Madison, 1978.

FARIA, Leandro Innocentini Lopes. **Prospecção tecnológica em materiais**: aumento da eficiência do tratamento bibliométrico. 2001. Tese (Doutorado em Ciência e Engenharia de materiais)-Universidade Federal de São Carlos, São Carlos, 2001.

FERNANDES, Rosali Pacheco. **Análises bibliométricas da produção científica dos grupos de pesquisa sobre física do estado sólido na América Latina**. 1973. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Bibliografia e Documentação, Rio de Janeiro, 1973.

FIGUEIREDO, Laura Maia de. **Distribuição da literatura geológica brasileira**: estudo bibliométrico. 1972. 32 p. Dissertação (Mestrado em Ciência da Informação)-

Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Bibliografia e Documentação, Rio de Janeiro, 1972.

FONSECA JÚNIOR, Wilson Corrêa da. **A comunicação organizacional: um estudo epistemológico.** 2007. Tese (Doutorado em Comunicação Social)_Universidade Metodista de São Paulo, São Paulo, 2007.

FORESTI, Noris Almeida Bethonico. **Estudo da contribuição das revistas brasileiras de Biblioteconomia e Ciência da Informação enquanto fonte de referência para a pesquisa.** 1989. 209 p. Dissertação (Mestrado em Ciência da Informação)-Universidade de Brasília, Brasília, 1989.

GOMES, Samuel. **Distribuição da literatura científica brasileira: estudo bibliométrico.** 1977. 53 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1977.

GUSMAO, Heloisa Rios. **Análise da literatura brasileira de siderurgia.** 1977. 58 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Bibliografia e Documentação, Rio de Janeiro, 1977.

HERNANDEZ CANADAS, Patrícia Liset. **Os periodicos "Ciencia Hoje" e "Ciencia e Cultura" e a divulgação da ciência no Brasil.** 1987. 190 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1987.

HIRATA, Geni Rodrigues da Costa. **Do texto científico ao texto de divulgação: um estudo quantitativo exploratório.** 1994. 175 p. Dissertação (Mestrado em Ciência da Informação)- Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1994.

IPPOLITO, Celina. **Análise comparativa da aquisição e circulação de periódicos em bibliotecas da Universidade de São Paulo na área médica e afim: uma metodologia bibliométrica.** 1973. 164 p. Dissertação (Mestrado em Ciências da Comunicação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Bibliografia e Documentação, Rio de Janeiro, 1973.

JOB, Ivone. **Educação Física no PPGMH/UFRGS: uma visão a partir da análise de citações e perfil dos pesquisadores.** 2006. 136 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal de Minas Gerais, Belo Horizonte, 2006.

LEAL, Igor Campos. **Análise de citações da produção científica de uma comunidade: a construção de uma ferramenta e sua aplicação em um acervo de teses e dissertações do PPGCI-UFMG.** 2005. 94 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

LOPES, Anna de Souza Ayres. **Estudo das citações da revista Pesquisa Agropecuária brasileira como instrumento para conhecimento do perfil dos autores.** 1992. 138 p. Dissertação (Mestrado em Ciência da Informação)-Universidade de Brasília, Brasília, 1992.

LOPEZ ROBLEDO, Edgar Leonel. **Estudo do desenvolvimento da literatura sobre física do estado sólido no México no período 1959-1971.** 1974. 188 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Bibliografia e Documentação, Rio de Janeiro, 1974

MAGALHÃES, Francyslene Abreu Costa. **Construção do saber no programa de doutorado em contabilidade no Brasil: plataformas teóricas e motivações.** 2006.

108 p. Dissertação (Mestrado em Ciências Contábeis)-Universidade de São Paulo, São Paulo, 2006.

MAIA, Elza Lima e Silva. **Comportamento bibliométrico da língua portuguesa, como veículo de representação da informação.** 1973. 57 p. Dissertação (Mestrado em Ciência da Informação)- Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Bibliografia e Documentação, Rio de Janeiro, 1973.

MAMFRIM, Flávia Pereira Braga. **Indexação automática derivativa em textos integrais em língua portuguesa.** 1990. 260 p. Dissertação (Mestrado em Ciência da Informação)- Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1990.

MARANHÃO, Ana Maria Neves. **Identidades semânticas analisadas bibliometricamente em textos bilíngües: estudo exploratório.** 1996. 159 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1996.

MARTINS, Helio Braga. **Produção científica dos professores do Centro de Geociências da UFPA: um estudo bibliométrico.** 2000. 104 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2000.

MELLO, Paula Maria Abrantes Cotta de. **A citação bibliográfica no contexto da comunicação científica: um estudo exploratório na área da Botânica.** 1990. 105 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1990.

MUGNAINI, Rogério. **Caminhos para adequação da avaliação da produção científica brasileira: impacto nacional versus internacional.** 2006. 254 p. Tese (Doutorado em Cultura e Informação)-Universidade de São Paulo, São Paulo, 2006.

MUGNAINI, Rogério. **Indicadores bibliométricos da base de dados Pascal como fonte de informação da produção científica e tecnológica do Brasil.** 2003. 133 p. Dissertação (Mestrado em Ciência da Informação)-Pontifícia Universidade Católica de Campinas, Campinas, 2003.

OBERHOFER, Cecilia Malizia Alves. **Uso da informação científica: obsolescência do objeto vs. obsolescência do conteúdo.** 1989. 164 p. Dissertação (Mestrado em Ciências da Comunicação)-Universidade de São Paulo, São Paulo, 1989.

OLIVEIRA, Margarida Pinto. **Estudo bibliométrico da literatura brasileira de esquistossomose.** 1975. 78 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Bibliografia e Documentação, Rio de Janeiro, 1975.

OLIVEIRA, Silas Marques de. **Comportamento da literatura brasileira de teologia adventista: análise bibliométrica.** 1981. 147 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal de Minas Gerais, Belo Horizonte, 1981.

PINTO, Aloisio de Arruda. **International savanna literature: a comparison of sources relevant to Brazilian researchers.** 1981. 98 p. Dissertação (Mestrado em Ciência da Informação)- University of Denver, Denver, 1981.

PITTELLA, Mônica Cardoso. **Análise de citação dos periódicos brasileiros de Biblioteconomia, 1972-1982.** 1990. 167 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal de Minas Gerais, Belo Horizonte, 1990.

QUEMEL, Maria Angélica Rodrigues. **Cruz e Sousa: do estudo bibliométrico ao estudo bibliográfico.** 1983. Dissertação (Mestrado em Ciências da Comunicação)-Universidade de São Paulo, São Paulo, 1983.

RAMOS, Virginia Maria Vasconcellos Prisco Paraíso. **Indexação automática derivativa de textos em língua portuguesa baseada em seus títulos e resumos: estudo exploratório.** 1992. 361 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1992.

RODRIGUEZ GARCIA, Marina Elvira del Socorro. **Química e químicos: estrutura da literatura e padrões de comunicação através da análise de citação da Revista Colombiana de Química.** 1988. 176 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1988.

ROLIM, Ivanilda Fernandes Costa. **A criação intelectual e artística no Brasil: uma análise da bibliografia nacional corrente, durante o período de 1951-1962.** 1976. 126 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Bibliografia e Documentação, Rio de Janeiro, 1976.

RUSSO, Mariza. **O padrão de comunicação na área de administração: o caso brasileiro e o caso americano.** 1991. 181. p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1991.

SAES, Sueli Gonzalez. **Estudo bibliométrico das publicações em economia da saúde, no Brasil, 1989-1998.** 2000 p. 115 p. Dissertação (Mestrado em Serviços de Saúde Pública)- Universidade de São Paulo, São Paulo, 2000.

SANTOS, Maria José Veloso da Costa. **A política brasileira de informação em geociência e tecnologia mineral: uma contribuição a sua implementação via análise bibliométrica da literatura.** 1992. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1992.

SILVA, Jussara Ferreira da. **Análise da disseminação e da divulgação científica na saúde coletiva do Instituto de Medicina Social da Universidade do Estado do Rio de Janeiro, tendo como instrumento a Lei de Zipf.** 1995. Dissertação (Mestrado em Ciência da Informação)- Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1995.

SILVA, Maurila Bentes de Mello e. **Estudo bibliométrico da produção científica do Núcleo de Altos Estudos Amazônicos (NAE) da Universidade Federal do Pará.** 2000. 159 p. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2000.

SOUZA, Alfredo A. C. Mendonça de. **Arqueologia brasileira (1975-1985): análise bibliométrica da literatura.** 1988. Dissertação (Mestrado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1988.

SOUZA, Luiz Antonio de. **Padrões de comunicação em saúde coletiva**: análise de citações das dissertações apresentadas do IMS-UERJ-1977-1992. 1995. Dissertação (Mestrado em Ciência da Informação)- Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1995.

VALOIS, Eliana Candeira. **Análise da produção técnico-científica dos pesquisadores do Programa Nacional de Pesquisa de Caprinos, no período de 1977 a 1988**. 1990. 214. p. Dissertação (Mestrado em Ciência da Informação)- Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1990.

ZIMBA, Horácio Francisco. **A pesquisa sobre medicina veterinária em Moçambique**: análise do desenvolvimento da produção científica 1947-2002. 2003. 114 p. Dissertação (Mestrado em Ciência da Informação)-Universidade de Brasília, Brasília, 2003.

ABSTRACT

Standards of information organization are imperative in bibliometric research, so that documents with similar characteristics may be efficiently collected from primary and secondary sources. Standards are also needed in order to improve the reliability of the data analysed. In order to identify the difficulties encountered by researchers within the field of information organization, we examined 59 reports of thesis and dissertations, as well as conducted a series of interviews with area experts. In general, the main difficulties were found in the standardization of journal titles and author names (physical description), and in the diversity of subject representation (subject description). The need to literature control by access to regularly updated databases was emphasized by most experts as one of the main difficulties for the conduction of research in the field of bibliometrics in Brazil

KEYWORDS: Bibliometry. Information Organization

Originals recebidos em: 08/02/2008

Texto aprovado em: 13/03/2008